Jour. Agric. Res. China (中華農業研究) 35(4) : 460~474 (1986)

Variance Component Analysis of Unbalanced Diallel Crosses[1]

Ching Liu[2]

Abstract : This paper presents an efficient and stable method for estimating variance components from an unbalanced diallel crosses. The method follows the general procedure for computing the sums of squares in a random model by formulating testable hypotheses as if all effects were fixed and then compute the corresponding sums of squares. Expected sums of squares are derived from the expectations of quadratic forms of the observations. Estimable functions for effects in the model are generated by using the Cholesky decomposition of the reduced X'X matrix on the grounds that the sequential reduction is suitable in obtaining the sums of squares of general combining abilities, specific combining abilities, maternal effects and reciprocal effects in an unbalanced diallel crosses. As far as the sampling properties of the estimated variance components are concerned, no other method has yet been proved to be more optimal uniformly than the present one. Same computational method can also be used to handle the mixed diallel model with a little modification in the expected values of the sums of squares. Computing covariance components can be done virtually the same way as computing the variance components. Numerical examples are given to illustrate the computational procedurs.

Introduction

Any set of matings among p individuals, strains or inbred lines which can be arranged into sets of symmetrical two-way classifications by parentage may be called a diallel crosses. Many and seemingly different diallel analyses have been presented in the literature. Diallel analyses differ in four main ways - in the sampling assumptions of the experimental materials, in the postulated underlying genetic mechanism, in the methods of estimation, and in the parental heterozygosity.

Kempthorne (1956) has discussed these methods of analysis in terms of variances of inbred parents, crossbred offspring and the covariance between parents and offspring. Griffing (1956) has classified the four different diallel crossing methods from which at least a p(p-1)/2 set of crosses can be obtained. He also discussed the analysis of

1. Contribution No. 1303 from Taiwan Agricultural Research Institute.

2. Senior Statistician, Department of Agronomy, TARI, Wufeng, Taichung Hsien, Taiwan 41301, R.O. C.

each in terms of the variance of general and specific combining abilities under two alternative assumptions with regard to the sampling nature of the experimental material. Theoretically, the important difference among these four methods depends upon whether not selfings are included. Griffing (1956) suggested that the term "diallel" should be used for those methods which include the selfings and the term "modified diallel" be used for those methods which do not. It was shown that with the modified diallel methods, unbiased estimates of the population variances can be obtained from mean squares of the appropriate analysis of variance (Griffing, 1958). Kempthorne (1956) derived the expectation of mean squares for a complete homozygous diallel. He showed that these mean squares alone could not yeild unbiased estimates of the desired parameters. Griffing (1958) also reached the similar conclusion for other diallel methods.

Maternal and recpirocal (sex-linkage) effects among the progenies of a set of homozygous diallel crosses has been examined by Wearden(1964). He derived the expected mean squares for three different analyses under two different models, assuming in each case either maternal or reciprocal effects. Topham (1966) has examined the implications of both effects occuring in the same material under the fixed model. Cockerham and Weir (1977) have given a quadratic analysis of reciprocal crosses of inbred lines for the balanced factorial and diallel mating designs.

Estimation of variance components in a diallel crosses is a special case of general problem of estimating variance components. Searle (1971) has given a good review on the topics in variance component estimation. Hartley's (1967) method of "synthesis" was one of the first methods developed for computing the coefficients in the expected values of mean squares for the mixed model. Gaylor et al. (1970) described the computation of expected values arising from the set of mean squares of the Forward Doolittle. Speed and Hocking (1974) developed the most computationally simple techniques to date and Goodnight and Speed (1980) modified the Speed and Hocking's technique to improve the computational efficiency.

Unbalanced data are those in which the numbers of observations in the subclasses of the model are not all the same, including cases where there are no observations in some subclasses. In contrast, balanced data are those in which the numbers of observations in all subclasses of the model are equal. Unequal subclass numbers always cause the different classes of effects to be non-orthogonal. This means that the different kinds of effects, such as general combining abilities and specific combining abilities, cannot be seperated directly without entanglement. In order to free these effects from confounding, it is necessary to consider all of the effects simultaneously. With equal subclass numbers, the effects are all mutually orthogonal and the statistics obtained directly from the class or subclass totals are equivalent to those obtained by the simultaneous consideration of all effects.

Diallel analysis with unbalanced data have not been discussed in detail before. In only one case (Harvey, 1960) have attempts been made to analyze diallel crosses data in the unbalanced case, without missing subclasses. The method of estimating variance components from an unbalanced diallel crosses will be discussed in this paper.

The Model

A diallel crosses with unbalanced data will be considered for estimating the variance components. There will be n_{ij} ($n_{ij} \geq 0$) observations of each cross in a completely randomized environmental design. More complicated environmental designs can be readily accomodated.

The model for the phenotypic value may include maternal, reciprocal (or sex-linkage) and environmental effects as well as the genotypic value which is usually composed of the general combining and specific combining abilities. The model can therefore be written as:

$$y_{ijk}=\mu+g_i+g_j+s_{ij}+m_i+r_{ij}+e_{ijk}, \quad (i \neq j) \quad \cdots\cdots [1]$$

where

μ = population mean of the crossbreds,

$g_i(g_j)$ = general combining ability effect for the i^{th} (j^{th}) parent,

s_{ij} = specific combining ability effect for the cross between the i^{th} and j^{th} parents such that $s_{ij} = s_{ji}$,

m_i = maternal effect of the i^{th} parent,

r_{ij} = reciprocal (sex-linkage) effect involving the reciprocal crosses between the i^{th} and j^{th} parents such that $r_{ij} = -r_{ji}$,

e_{ijk} = environmental effect associated with the ijk^{th} individual observation of the crossbreds.

Under the assumptions of the model (Random Model), all terms (except μ) are taken as being normally distributed with zero means and their respective variances.

Following Searle (1971), [1] can be represented by a general random model as:

$$Y=\mu 1+\sum_{i=1}^{k} X_i\beta_i+e \quad \cdots\cdots [2]$$

$$=X\beta+e$$

where

(1) Y is an n-vector of observations,

(2) 1 is an n-vector of 1's, $X_1, \cdots, X_k$ are nxm_i known matrices and $X=[1 \mid X_1 \mid \cdots \mid X_k]$,

(3) μ is general mean,

(4) The vectors $\beta_i (i=1, \cdots, k)$ are assumed independent of each other and are distributed $N(0, \sigma_i^2 I_{mi})$,

(5) e is an n-vector of random error and are assumed $N(0, \sigma_e^2 I)$.

On making the above normality assumptions.

$E(Y)=\mu 1$,

$$Var(Y)=\sum_{i=1}^{k} X_i X_i'\sigma_i^2+I\sigma_e^2.$$

Computing Sums of Squares

The qeneral procedure for computing the sums of squares (SS) in a random model is to formulate testable hypotheses as if all effects were fixed and then compute the corresponding SS.

Letting (1) $\beta=\begin{pmatrix}\mu\\ \beta_1\\ \vdots\\ \beta_k\end{pmatrix}$

(2) $(X'X)^-$=any generalized inverse of $(X'X)$

(3) $b=(X'X)^-(X'Y)$

(4) L=any matrix of full row rank in the row space of X,

then

$$SS_L=SS(H_0: L\beta=0)$$
$$=(Lb)'(L(X'X)^-L')^{-1}(Lb)$$
$$=Y'Q_LY \quad \cdots\cdots [3]$$

where $Q_L=X(X'X)^-L'(L(X'X)^-L')^{-1}L(X'X)^-X'$.

Expected Values of Sums of Squares

Equation [3] is a quadratic form in Y and any symmetric quadratic form Y'QY has (Searle, 1971) :

$$E(Y'QY)=\mu^2 1'Q1+\sum_{i=1}^{k} tr\,(X_i'QX_i)\sigma_i^2+tr(Q)\sigma_e^2 \quad \cdots\cdots [4]$$

Letting $X=[1 \mid X_1 \mid \cdots \mid X_k]$, all of the matrices except Q needed in [4] are submatrices of X'QX. It is easily seen that, from [4], only the elements of $X'Q_LX$ are needed to evaluate the expected value of [3] if [2] is assumed to be the true model.

Since L is in the row space of X, from the theory of estimability

$$L=L(X'X)^-X'X$$

and therefore

$$X'Q_LX=L'(L(X'X)^-L')^{-1}L.$$

Let U=Cholesky decomposition (upper) of $L(X'X)^-L'$,

(i.e., $L(X'X)^-L'=U'U$)

then $X'Q_LX=L'(L(X'X)^-L')^{-1}L$

$$=L'(U'U)^{-1}L$$
$$=C'C \text{ , where } C=(U')^{-1}L$$
$$=\begin{pmatrix} C_0'C_0 & C_0'C_1 & \cdots & C_0'C_k\\ C_1'C_0 & C_1'C_1 & \cdots & C'_1C_k\\ \vdots & \vdots & \vdots & \vdots\\ C_k'C_0 & C_k'C_1 & \cdots & C_k'C_k\end{pmatrix}$$

From [4], we have

$$E(SS_L)=C_0^2\mu^2+\sum_{i=1}^{k} tr\,(C_i'C_i)\,\sigma_i^2+n_L\sigma_e^2$$

$$=C_0^2\mu^2+\sum_{i=1}^{k} SSQ(C_i)\sigma_i^2+n_L\sigma_e^2 \quad \cdots\cdots [5]$$

where SSQ (C_i) = sum of squares of the elements of the C_i submatrix,
and n_L = the number of rows in L.

It is seen from [5] that C'C need not be computed and the matrix $C=(U')^{-1}L=[C_0 \mid C_1 \mid \cdots \mid C_k]$ of the same dimensions of L is all that is needed for computing the expected values of SS_L.

Reducing Dimensionality

Let the model in [2] be written as

$$Y=X\beta+e$$

$$=[X_1 \mid X_2]\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}+e$$

where X_1 has full column rank and $X_2=X_1M$ for some matrix M.
Imposing $\hat{\beta}_2=0$, we have

$$X'X=\left(\begin{array}{c|c} X_1'X_1 & X_1'X_2 \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right),\quad (X'X)^-=\left(\begin{array}{c|c} (X_1'X_1)^{-1} & 0 \\ \hline 0 & 0 \end{array}\right)$$

where $X_1'X_1$ has full row rank , equal to r(X), and $(X'X)^-$ is a g2-inverse of X'X.

Since $L(X'X)^-L'=[L_1 \mid L_2]\left(\begin{array}{c|c} (X_1'X_1)^{-1} & 0 \\ \hline 0 & 0 \end{array}\right)\begin{pmatrix}L_1'\\ L_2'\end{pmatrix}$

$$=L_1(X_1'X_1)^{-1}L_1'$$
$$=U'U, \text{ (Cholesky decomposition)}$$

then $C=(U')^{-1}L=[C_0 \mid C_1 \mid \cdots \mid C_k]$
Similarly,

$$SS_L=(Lb)'(L(X'X)^-L')^{-1}(Lb)$$
$$=(L_1b_1)'(L_1(X_1'X_1)^{-1}L_1')^{-1}(L_1b_1)$$

where $b=(X'X)^-(X'Y)=\begin{pmatrix}(X_1'X_1)^{-1}(X_1'Y)\\ 0\end{pmatrix}=\begin{pmatrix}b_1\\ 0\end{pmatrix}$

and $E(SS_L)=C_0^2\mu^2+\sum_{i=1}^{k} SSQ(C_i)\sigma_i^2+n_L\sigma_e^2$

Thus without affecting anything, imposing restrictions reduces the computational burden significantly.

Generating Estimable Functions

In order for a hypothesis H_0: $L\beta=0$ to be testable, $L\beta$ must first be estimable. For any L we say that $L\beta$ is estimable if and only if a linear combination of the Y's exists which has an expected value of $L\beta$. Since any linear combination of the Y's, say KY, has $E(KY)=KX\beta$, then for $L\beta$ to be estimable, L must be a linear combination of the rows of the X matrix. In other words, L must be in the row space of X. This leads to the definition that: $L\beta$ is estimable if and only if there exists a K such

that L=KX (Goodnight, 1978).

It is clear from the above definition of estimability that if an L is to be constructed such that $L\beta$ is estimable, then only linear combinations of the rows of X need be considered. In fact any linear combination of the rows of X yields an L such that $L\beta$ is estimable. Thus the X matrix or any matrix (with the same row rank of X) constructed from row operations on X may be used as a generating set for constructing any and all L's. Obviously an infinite number of L's can be generated. For a given model, an entire set of L's must be generated, one L for each effect in the model.

As far as the estimation of variance components in a diallel crosses is concerned, the simplest sets of L's is generated by computing the Cholesky decomposition matrix of X'X, and letting the L's for each effect be the non-zero rows associated with that effect.

Cholesky decomposition matrix of full rank matrix X'_1X_1 is obtained most conveniently by first computing the Forward Doolittle matrix from $X_1'X_1$:

$$X_1'X_1 \xrightarrow{\text{Forward Doolittle}} A_1, \text{ (upper triangular matrix).}$$

If each row of A_1 is divided by the square root of its diagonal element to produce a matrix Z_1, then Z_1 is the Cholesky decomposition matrix of $X_1'X_1$ with $X'_1X_1=Z_1'Z_1$.
For symmetric positive semi-definite matrix X'X,

$$\text{let } X'X=\left(\begin{array}{c|c} X'_1X_1 & X'_1X_2 \\ \hline X'_2X_1 & X'_2X_2 \end{array}\right) \text{ and } X_1'X_1=Z_1'Z_1$$

where $X_1'X_1$ has full rank.

$$\text{If } Z=\left(\begin{array}{c|c} Z_1 & (Z_1')^{-1}(X_1'X_2) \\ \hline 0 & 0 \end{array}\right), \text{ then } X'X=Z'Z.$$

(i.e., Z is the Cholesky decomposition matrix of X'X.)
Therefore $[Z_1 \mid (Z'_1)^{-1}(X'_1X_2)]$ can be used to generate the entire set of L's, one L for each effect in the model.

$$\text{If } L=[Z_1 \mid (Z'_1)^{-1}(X'_1X_2)] = \left(\begin{array}{c} L_0 \\ \hline L_1 \\ \hline \vdots \\ \hline L_s \\ \hline \vdots \\ \hline L_k \end{array}\right) = \left(\begin{array}{c|c} L_{01} & L_{02} \\ \hline L_{11} & L_{12} \\ \hline \vdots & \vdots \\ \hline L_{s1} & L_{s2} \\ \hline \vdots & \vdots \\ \hline L_{k1} & L_{k2} \end{array}\right)$$

then (1) $L_s(X'X)^-L_s'=L_{s1}(X_1'X_1)^{-1}L'_{s1}=I$ (identity matrix)

(2) $C_s=L_s=[C_{s0} \mid C_{s1} \mid \cdots \mid C_{sk}]$

(3) $SS_{Ls}=SS(H_0 : L_s\beta=0)=(L_sb)'(L_s(X'X)^-L'_s)^{-1}(L_sb)$

$=(L_{s1}b_1)'(L_{s1}(X_1'X_1)^{-1}L'_{s1})^{-1}(L_{s1}b_1)$

$=(L_{s1}b_1)'(L_{s1}b_1)$ ..... [6]

where $b=\left(\begin{array}{c} b_1 \\ \hline 0 \end{array}\right)=\left(\begin{array}{c} (X_1'X_1)^{-1}(X'_1Y) \\ \hline 0 \end{array}\right)$

(4) $E(SS_{Ls}) = C^2_{s0}\mu^2+\sum_{i=1}^{k} SSQ(C_{si})\sigma_i^2+n_{Ls}\,\sigma_e^2$ ..... [7]

The computational burden is further reduced if the Cholesky decomposition matrix of X'X is used for generating the L's.

Numerical Examples

The data given in Table 1 were used to illustrate the computational procedures developed in this paper. Suppose the data represent average daily gains of the progenies of a 4×4 diallel crosses in an animal breeding program. Data set 1 is balanced with 4 progenies for each cross. The other data sets are all unbalanced (unequal numbers of progenies for each cross), but with different structures to include some missing-subclass (cross) cases. Data set 2 has every cross present. Data set 3 has a cross (2×4) missing. Data set 4 has a cross (2×4) and a corresponding reciprocal cross (4×2) missing. Data set 5 has 2 crosses (2×4 and 2×1) of the same female parent missing.

Table 1. Data usde in numerical examples

Cross (F×M)	Data set						
	1	2	3	4	5	6	7
1×2	114	114	114	114	114	114	114
	119	119	119	119	119	119	119
	132	132	132	132	132	132	132
	148	148	148	148	148	148	148
1×3	126	126	126	126	126	126	126
	125	125	125	125	125	125	125
	113	113	113	113	113	113	113
	119	—	—	—	—	—	—
1×4	104	104	104	104	104	104	104
	108	108	108	108	108	108	108
	104	104	104	104	104	104	104
	100	100	100	100	100	100	100
2×1	127	127	127	127	—	127	—
	128	128	128	128	—	128	—
	114	—	—	—	—	—	—
	122	—	—	—	—	—	—
2×3	137	137	137	137	137	137	137
	135	135	135	135	135	135	135
	126	126	126	126	126	126	126
	131	131	131	131	131	131	131

Table 1. (續)

Cross (F×M)	Data set						
	1	2	3	4	5	6	7
2×4	115	115	—	—	—	—	—
	101	101	—	—	—	—	—
	101	101	—	—	—	—	—
	112	—	—	—	—	—	—
3×1	112	112	112	112	112	112	112
	133	133	133	133	133	133	133
	113	113	113	113	113	113	113
	124	—	—	—	—	—	—
3×2	134	134	134	134	134	134	134
	138	138	138	138	138	138	138
	129	129	129	129	129	129	129
	128	128	128	128	128	128	128
3×4	112	112	112	112	112	—	—
	111	111	111	111	111	—	—
	97	—	—	—	—	—	—
	107	—	—	—	—	—	—
4×1	100	100	100	100	100	100	100
	114	114	114	114	114	114	114
	94	94	94	94	94	94	94
	117	117	117	117	117	117	117
4×2	116	116	116	—	116	116	116
	114	114	114	—	114	114	114
	106	106	106	—	106	106	106
	110	—	—	—	—	—	—
4×3	97	97	97	97	97	97	97
	98	98	98	98	98	98	98
	103	103	103	103	103	103	103
	96	96	96	96	96	96	96
No. of obs.	48	40	37	34	35	35	33

Data set 6 has 2 crosses (2×4 and 3×4) of the same male parent missing. Data set 7 has 3 crosses (2×4, 2×1 and 3×4) missing, of which 2 crosses have the same female parent and 2 crosses have the same male parent. [Table 1]

For model in [1], the reduced X'X matrix for generating estimable functions and the estimable functions generated by the Cholesky decomposition of the reduced X'X matrix were given in Table 2 and Table 3, respectively, for data set 1 only to save space. Values for other data sets were obtained in a similar manner. Restrictions on general combining abilities (gca) and on maternal effects (me) were imposed by the usual way. Restrictions on specific combining abilities (sca) were applied on the last effect of every female parent first, then applied the remaining constraint on effect

Table 2. Reduced X'X matrix of data set 1 for generating estimable functions

Effect	Source of variation											
	mean(μ)	gca(g)			sca(s)		me(m)			re(r)		
μ	48	24	24	24	8	8	12	12	12	0	0	0
g_1	24	24	8	8	8	8	12	4	4	0	0	0
g_2	24	8	24	8	8	0	4	12	4	0	0	0
g_3	24	8	8	24	8	8	4	4	12	0	0	0
g_4	24	8	8	8	0	0	4	4	4	0	0	0
s_{12}	8	8	8	0	8	0	4	4	0	0	0	0
s_{13}	8	8	0	8	0	8	4	0	4	0	0	0
s_{14}	8	8	0	0	0	0	4	0	0	0	0	0
s_{23}	8	0	8	8	0	0	0	4	4	0	0	0
s_{24}	8	0	8	0	0	0	0	4	0	0	0	0
s_{34}	8	0	0	8	0	0	0	0	4	0	0	0
m_1	12	12	4	4	4	4	12	0	0	4	4	0
m_2	12	4	12	4	4	0	0	12	0	-4	0	4
m_3	12	4	4	12	0	4	0	0	12	0	-4	-4
m_4	12	4	4	4	0	0	0	0	0	0	0	0
r_{12}	0	0	0	0	0	0	4	-4	0	8	0	0
r_{13}	0	0	0	0	0	0	4	0	-4	0	8	0
r_{14}	0	0	0	0	0	0	4	0	0	0	0	0
r_{23}	0	0	0	0	0	0	0	4	-4	0	0	8
r_{24}	0	0	0	0	0	0	0	4	0	0	0	0
r_{34}	0	0	0	0	0	0	0	0	4	0	0	0

which have the highest i+j value (i and j stand for the female and male parent indices, respectively). Restrictions on reciprocal effects (re) were applied on the last effect of every male parent first, then applied the remaining constraint on effect which have the highest i+j value . For data with i×j cross missing, an additional constraint must be applied on the i×j reciprocal effect, simply because this reciprocal effect is no longer estimable. For data with both i×j and j×i crosses missing, two additional constraints must be employed. One for the i×j specific combining ability and the other for the i×j reciprocal effect, since both these two effects are no longer estimable.

Table 3. Estimable functions of data set 1 generated by Cholesky decomposition of reduced X'X matrix ($\times 10^{-3}$)

Effect	Source of variation											
	mean(μ)	gca(g)			sca(s)		me(m)			re(r)		
μ	6928	0	0	0	0	0	0	0	0	0	0	0
g_1	3464	3464	0	0	0	0	0	0	0	0	0	0
g_2	3464	-1155	3266	0	0	0	0	0	0	0	0	0
g_3	3464	-1155	-1633	2828	0	0	0	0	0	0	0	0
g_4	3464	-1155	-1633	-2828	0	0	0	0	0	0	0	0
s_{12}	1155	1155	1633	0	1633	0	0	0	0	0	0	0
s_{13}	1155	1155	- 817	1414	-817	1414	0	0	0	0	0	0
s_{14}	1155	1155	- 817	-1414	-817	-1414	0	0	0	0	0	0
s_{23}	1155	-1155	817	1414	-817	-1414	0	0	0	0	0	0
s_{24}	1155	-1155	817	-1414	-817	1414	0	0	0	0	0	0
s_{34}	1155	-1155	-1633	0	1633	0	0	0	0	0	0	0
m_1	1732	1732	0	0	0	0	2450	0	0	0	0	0
m_2	1732	- 577	1633	0	0	0	-817	2309	0	0	0	0
m_3	1732	- 577	- 817	1414	0	0	-817	-1155	2000	0	0	0
m_4	1732	- 577	- 817	-1414	0	0	-817	-1155	-2000	0	0	0
r_{12}	0	0	0	0	0	0	1633	-1155	0	2000	0	0
r_{13}	0	0	0	0	0	0	1633	577	-1000	-1000	1732	0
r_{14}	0	0	0	0	0	0	1633	577	1000	-1000	-1732	0
r_{23}	0	0	0	0	0	0	0	1732	-1000	1000	- 577	1633
r_{24}	0	0	0	0	0	0	0	1732	1000	1000	577	-1633
r_{34}	0	0	0	0	0	0	0	0	2000	0	1155	1633

From the estimable functions generated by the Cholesky decomposition of the reduced X'X matrix, sums of squares (SS) and the expected values of SS for general combining abilities, specific combining abilities, maternal effects and reciprocal effects were easily computed from [6] and [7], respectively, and were listed in Table 4. Estimated values of varianae components were summarized in Table 5. Some estimates were negative, but this was largely due to the sampling error of the data itself and

Table 4. Sums of squares (SS) and expected values of SS for estimable functions generated by Cholesky decomposition of reduced X'X matrix

S. O. V.	Data set	DF	SS(H_0:Lβ=0)	Coefficients in Exp(SS)					
				μ^2	σ_g^2	σ_s^2	σ_m^2	σ_r^2	σ_e^2
mean(μ)	1	1	644960.33344	48	48.00000	8.00000	12.00000	0.00000	1
	2	1	542657.02494	40	40.00000	6.80000	10.10000	0.20000	1
	3	1	509539.56744	37	37.24324	6.62162	9.70270	0.45946	1
	4	1	472001.05876	34	35.05882	6.94118	8.88235	0.23529	1
	5	1	477244.82846	35	35.37143	6.42857	9.68571	0.82857	1
	6	1	484747.45762	35	35.37143	6.42857	9.34286	0.82857	1
	7	1	452439.27307	33	33.27273	6.21212	9.30303	...	1
gca(g)	1	3	5542.37500	0	48.00000	24.00000	12.00000	0.00000	3
	2	3	5239.89583	0	40.00000	20.00000	10.33333	0.66667	3
	3	3	4974.73876	0	36.75675	18.82282	10.18619	1.87387	3
	4	3	4889.03939	0	32.94118	20.20168	8.59384	0.71709	3
	5	3	4768.53881	0	34.62857	17.90476	10.31429	3.40952	3
	6	3	5030.02137	0	34.62857	17.90476	10.65714	3.40952	3
	7	3	4851.37311	0	32.72727	16.78788	10.94697	...	3
sca(s)	1	2	191.29167	0	0.00000	16.00000	0.00000	0.00000	2
	2	2	134.70417	0	0.00000	13.20000	0.23333	0.46667	2
	3	2	81.15201	0	0.00000	11.55556	0.77778	2.00000	2
	4	1	73.36012	0	0.00000	6.85714	0.19048	0.38095	1
	5	2	70.34094	0	0.00000	10.66667	1.33333	3.42857	2
	6	2	178.81316	0	0.00000	10.66667	1.33333	3.42857	2
	7	2	153.89583	0	0.00000	10.00000	1.75000	...	2
me(m)	1	3	87.18750	0	0.00000	0.00000	24.00000	24.00000	3
	2	3	48.58302	0	0.00000	0.00000	19.33333	19.76246	3
	3	3	73.56640	0	0.00000	0.00000	16.33333	20.07312	3
	4	3	73.56640	0	0.00000	0.00000	16.33333	20.07312	3
	5	3	86.20833	0	0.00000	0.00000	13.66667	21.07246	3
	6	3	16.05978	0	0.00000	0.00000	13.66667	21.07246	3
	7	3	16.12500	0	0.00000	0.00000	11.00000	...	3
re(r)	1	3	155.81250	0	0.00000	0.00000	0.00000	24.00000	3
	2	3	253.79198	0	0.00000	0.00000	0.00000	18.90420	3
	3	2	168.64194	0	0.00000	0.00000	0.00000	12.59355	2
	4	2	168.64194	0	0.00000	0.00000	0.00000	12.59355	2
	5	1	155.25000	0	0.00000	0.00000	0.00000	6.26087	1
	6	1	0.81522	0	0.00000	0.00000	0.00000	6.26087	1
	7	–	...	–	...	...	...	...	–
error(e)	1	36	2143.00000	0	0.00000	0.00000	0.00000	0.00000	36
	2	28	1827.00000	0	0.00000	0.00000	0.00000	0.00000	28
	3	26	1696.33333	0	0.00000	0.00000	0.00000	0.00000	26
	4	24	1640.33333	0	0.00000	0.00000	0.00000	0.00000	24
	5	25	1695.83333	0	0.00000	0.00000	0.00000	0.00000	25
	6	25	1695.83333	0	0.00000	0.00000	0.00000	0.00000	25
	7	24	1695.33333	0	0.00000	0.00000	0.00000	0.00000	24

had not much to do with the computational method employed. Negative estimates occurred in even the balanced data (Table 5).

Table 5. Estimated values of μ^2 and variance components

Data set	Estimated values of					
	μ^2	σ_g^2	σ_s^2	σ_m^2	σ_r^2	σ_e^2
1	13325.19271	110.20312	4.51476	− 2.85938	− 0.94878	59.52778
2	13438.79583	128.62971	0.39998	−10.75056	3.07032	65.25000
3	13637.45220	134.76743	−3.51535	−11.20286	3.02971	65.24359
4	13734.09118	144.49875	0.90032	−11.16717	2.53681	68.34722
5	13498.40773	142.96354	−6.84836	−30.11073	13.96238	67.83333
6	13708.86010	135.88499	7.13694	2.78966	−10.70428	67.83333
7	13565.85548	145.27819	4.37667	−17.79924	···	70.63889

Although data used in this paper vary considerably in degree of unbalancedness, the estimated values of variance components for each effect are fairly consistent among 7 data sets. This demonstrates the stable nature of the computing algorithm developed in this paper.

Discussion

Estimating variance components from unbalanced data is not as straightforward as from balanced data. Several methods of estimation are available. Most of these methods reduced to the analysis of variance method for balanced data. Only a few comparative studies have been made on the relative merits of the different methods (Bush and Anderson, 1963; Anderson and Crump, 1967 and Kussmaul and Anderson, 1967). None of the methods has yet been clearly established as superior to the others. All of the methods involve relative cumbersome algebra. Discussion of unbalanced data can therefore easily deteriorate into a "welter of symbols" (Searle, 1971).

Methods 1 and 3 of Henderson (1953) and complete least-squares analysis are frequently used for estimating variance components from unbalanced data under the random model. All of these methods use, in one way or another, quadratic forms of the observations. The general methodology is to obtain expected values of quadratic forms and to equate them to their observed values. This is exactly what is done with sums of squares, which are quadratic forms of the observations, in the analysis of variance method for balance data. Whereas with balanced data their is obviously only one set of quadratic forms, they lead to estimators that are unbiased and have minimum variance properties. There are many sets of quadratic forms that can be used for unbalanced data. The simplest way to evaluate these estimators would be to compare their sampling variances , but to compare the sampling variances of different estimators is difficult because the variances themselves involve the true variance components even if

normality is assumed. It should be noted that although the change from balanced data to unbalanced data might appear relatively small on the surface, it causes enormous change in the tractability of methods and the properties of estimators. Comparative studies of different estimators arising from different quadratic forms, analytically or numerically, are therefore not simple because of the algebric complexity of unbalanced data involved. No particular set of quadratic forms has yet been established as more optimal than any other set. The only universal outcome of these different methods is that they all provide unbiased estimators. Unlike testing fixed effects where appropriate F-statistic is a problem, unbiased variance components can always be estimated by equating the expected values of quadratic forms to their observed values. Therefore an efficient and stable method for computing the observed values of quadratic forms and their expected values is most desirable.

Variance component estimators derived by the method presented in this paper are linear combinations of quadratic forms of the observations. Each quadratic form can be expressed in the form Y'QY for an appropriate Q-matrix and hence its sampling variance can be derived from application of $\mathrm{Var}(Y'QY)=2\ \mathrm{tr}(QV)+4\mu^2 1'QVQ1$, where $V=\mathrm{Var}(Y)=\sum_{i=1}^{k} X_iX_i'\sigma_i^2+I\sigma_e^2$, and the sampling covariance between any two quadratic forms can be derived from $\mathrm{Cov}(Y'QY, Y'RY)=2\ \mathrm{tr}(QVRV)+4\mu^2 1'QVR1$. In this way the sampling variance of linear combinations of these quadratic forms can be obtained, although the details are somewhat lengthy, involving considerable matrix manipulation.

The use of the Cholesky decomposition matrix of X'X in generating the estimable functions for effects in an unbalanced diallel crosses, as used in this paper, is proven to be both efficient and stable as well as appropriate. Methods 1 and 3 of Henderson and complete least-squares analysis, unreparametrized or reparametrized, should not be used in design matrix structure as unbalanced diallel. This is because in an unbalanced diallel, general combining abilities should not be adjusted for specific combining abilities and maternal effects should not be adjusted for reciprocal effects. Therefore sequential reductions, which is the outcome of the use of the Cholesky decomposition matrix of X'X, is proper in obtaining the sums of squares of general combining abilities, specific combining abilities, maternal effects and reciprocal effects in an unbalanced diallel crosses.

Although the diallel crosses model used in this paper is assumed random, same computational method can also be used to handle the mixed model with a little modification in the expected values of the sums of squares. Besides, computing sums of cross-products from an unbalanced data is analogus to computing sums of squares, therefore computing covariance components can be done virtually the same way as computing variance components. As far as the sampling properties of the estimated variance components are concerned, no other method has yet been proved to be more optimal uniformly than the present one.

Literature cited

1. Anderson, R. L. and P. P. Crump. 1967. Comparisons of designs and estimation procedures for estimating parameters in a two-stage nested process. Technometrics 9 : 499-516.
2. Bush, N. and R. L. Anderson. 1963. A comparison of three different procedures for estimating variance components. Technometrics 5 : 421-440.
3. Cockerham, C. C. and B. S. Weir. 1977. Quadratic analyses of reciprocal crosses. Biometrics 33 : 187-203.
4. Gaylor, D. W., H. L. Lucas and R. L. Anderson. 1970. Calculation of expected mean squares by the abbreviated Doolittle methods. Biometrics 26 : 641-655.
5. Goodnight, J. H. 1978. Tests of hypotheses in fixed effects linear models. SAS Technical Report R-101.
6. Goodnight, J. H. and F. M. Speed. 1980. Computing expected mean squares. Biometrics 36 : 123-125.
7. Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. Aust. J. Biol. Sci. 9 : 463-493.
8. Griffing, B. 1958. Application of some sampling variables in the identification of methods which yield unbiased estimates of genotypic variance components. Aust. J. Biol. Sci. 11 : 219-245.
9. Hartley, H. O. 1967. Expectations, variances and covariances of ANOVA mean squares by 'synthesis'. Biometrics 23 : 105-114.
10. Harvey, W. R. 1960. Least-squares analysis of data with unequal sub-class numbers. U. S. D. A. Agricultural Research Service Publication No. ARS 20-8. 157 pp.
11. Henderson, C. R. 1953. Estimation of variance and covariance components. Biometrics 9 : 226-252.
12. Kempthorne, 0. 1956. The theory of the diallel cross. Genetics 41 : 451-459.
13. Kussmaul, K. and R. L. Anderson. 1967. Estimation of variance components in two stage nested designs with composite samples. Technometrics 9 : 373-390.
14. Searle, S. R. 1971. Topics in variance component estimation. Biometrics 27 : 1-76.
15. Speed, F. M. and R. R. Hocking. 1974. Computation of expectations, variances and covariances of ANOVA mean squares. Biometrics 30 : 157-169.
16. Topham, P. B. 1966. Diallel analysis involving maternal and maternal interaction effects. Heredity 21 : 665-674.
17. Wearden, S. 1964. Alternative analyses of the diallel cross. Heredity 19 : 669-680.

不均衡全互交之變方成分分析[1]

劉 清[2]

摘 要

本文提出一種有效且穩定的方法來估算不均衡全互交資料之變方成分。本法先將所有效應視爲固定型以構成可測擬說（testable hypothesis），後再計算逢機型模式中各項效應之平方和，平方和之期望值是由觀測值二次形（quadratic form）之期望值導出。模式中各項效應之可估函數（estimable function）是由減縮（reduced）X′X矩陣經Cholesky分解而得來，此因在不均衡全互交資料中，逐次減縮法（sequential reduction）最適合用來計算一般組合力、特殊組合力、母系效應及反交效應等之平方和。就估算所得之變方成分的取樣性質（sampling properties）而言，尚未有其它方法經證實較本法一致爲優。本文之計算方法亦可應用於混雜型全互交模式，僅需將平方和之期望值部份作些許修改。由於變方成分計算與變積成分計算之共同性，本法亦可用來估算變積成分。文中附有數例以說明計算過程。

1.臺灣省農業試驗所 研究報告第 1303 號。

2.本所農藝系研究員。臺灣省 臺中縣 霧峰鄉。